Development of a Real-time Objective Gas-Liquid Flow Regime Identifier using Kernel Methods

Edem Nsefik Eyo, Karl Ezra Salgado Pilario, Liyun Lao, and Gioia Falcone

***Abstract*—Currently, flow regime identification for closed channels have mainly been direct subjective methods. This presents a challenge when dealing with opaque test sections of the pipe or at gas-liquid flow rates where unclear regime transitions occur. In this paper, we develop a novel real-time objective flow regime identification tool using conductance data and kernel methods. Our experiments involve a flush mounted conductance probe that collects voltage signals across a closed channel. The channel geometry is a horizontal annulus, which is commonly found in many industries. Eight distinct flow regimes were observed at selected gas-liquid flow rate settings. An objective flow regime identifier was then trained by learning a mapping between the probability density function (PDF) of the voltage signals and the observed flow regimes via kernel principal components analysis (KPCA) and multi-class Support Vector Machine (SVM). The objective identifier was then applied in real-time by processing a moving time-window of voltage signals. Our approach has: (a) achieved more than 90% accuracy against visual observations by an expert for static test data; (b) successfully visualized conductance data in 2-dimensional space using virtual flow regime maps, which are useful for tracking flow regime transitions; and, (c) introduced an efficient real-time automatic flow regime identifier, with only conductance data as inputs.**

***Index Terms*—conductance, KPCA, non-invasive, regime chart, SVM, virtual flow regime map.**

I. Introduction

THE accurate determination of flow regimes is a fundamental problem in the operation of transport systems in a plant, e.g. piping, that require gas and liquid to flow through them simultaneously [1]. In these systems, the task of predicting flow properties such as pressure drop and liquid holdup is greatly dependent on the governing flow regime. Different researchers have indicated that flow regimes influence mass and momentum transfer during two-phase flow and as such, modeling strategies have been based on the analysis of the prevailing flow regimes. Also, modeling based on the identification of flow regimes can result in increased safety and performance in two-phase flow systems [2].

In this paper, an objective flow regime identification method was developed by using conductance data and kernel methods. Data sets of voltage signals were collected from conductance rings flush mounted on a horizontal annular channel, where 8 different flow regimes were observed. Our approach is to project the probability density function (PDF) features from the voltage signals onto a reduced space of 2 dimensions, via kernel principal components analysis (KPCA). This step effectively visualizes the conductance data in 2D space, without knowledge of the current gas and liquid flow rates. In the newly constructed map, boundaries between flow regimes are then established using a kernel support vector machine (SVM). This step assigns every point in the map to a flow regime. For real-time identification, a moving time-window of voltage signals is processed by the trained KPCA-SVM identifier so that an exact flow regime is determined at every time step. We introduce regime charts and virtual flow regime maps to communicate results to human operators.

As opposed to our approach, current flow regime identification methods are mostly subjective [1], [2]. Direct methods include visual observation of the flow through transparent test sections, such as the use of photography or high speed camera imagery [3]–[5]. However, these methods have limitations: (i) when dealing with channels which are opaque or hidden from view, or (ii) at too high flow rates of the phases where transition boundaries are unclear. On the other hand, indirect methods have been developed to deal with such limitations. These are based primarily on the statistical analysis of the fluctuating nature of the flow [6]. The common practice has been to undertake experiments which measure certain intrinsic flow parameters or features that can distinguish between flow regimes, from which an identifier can be trained [7].

Furthermore, gas-liquid flow in *annular* closed channels is common in several industrial processes including chemical, petroleum and nuclear systems. However, there are very limited studies on characterizing flow regimes in annuli, as compared to circular pipe flow. Originally, Kelessidis and Dukler [8] and Das *et al.* [9] employed probability density function (PDF) of void fraction as indicators of two-phase flow in vertical annuli. Many other studies have similarly reported the use of visual observations from high speed, digital and video cameras [10]–[13]. For instance, Eyo and Lao [13] recently presented systematic flow regime characterization in horizontal concentric and fully eccentric annuli using high speed camera images and PDF of instantaneous liquid holdup obtained using conductance probes. In all of these studies,

Manuscript received July 31, 2018; revised January 22, 2019; accepted April 03, 2019. This work was supported in part by the Petroleum Technology Development Fund (PTDF) of Nigeria and in part by the DOST-ERDT Faculty Development Fund of the Republic of the Philippines. (*Corresponding author: Liyun Lao*)

E. N. Eyo and L. Lao are with the School of Water, Energy, and Environment, Cranfield University, MK43 0AL, U.K. (email: edem-nsefik.eyo@cranfield.ac.uk; l.lao@cranfield.ac.uk)

K. E. S. Pilario is with the School of Water, Energy, and Environment, Cranfield University, MK43 0AL, U.K., on study leave from the Department of Chemical Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippines. (email: k.pilario@cranfield.ac.uk)

G. Falcone is with the School of Engineering, University of Glasgow, Glasgow, G12 8QQ, U.K. (email: Gioia.Falcone@glasgow.ac.uk)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier

flow regime identification was still prone to subjectivity of the investigators. The precise identification of flow regimes, especially at transition regions, remains a challenging task.

The current need to develop *objective* flow regime identification tools [2] is now being addressed by the use of pattern recognition or machine learning techniques. Yan *et al.* [14] has recently reported a survey of such techniques used in multiphase flow metering. Artificial neural networks (ANN) [5], [15]–[21], support vector machines (SVM) [22]–[24], fuzzy systems [25]–[27], and their variants are popularly used. Despite the prevalence of these techniques, it is worth noting that flow regime identifiability is still influenced by the chosen instrumentation for collecting data, the size and frequency of collected data, and the informative quality of features extracted from the data.

Non-invasive instrumentation is currently preferred for measuring two-phase flow features for subsequent pattern recognition. Abbagoni and Yeung [28], for instance, used frequency-domain features from ultrasonic Doppler signals with ANNs, giving 80-90% classification accuracy. Similarly, Julia *et al.* [7] studied adiabatic upward two-phase flow using bubble chord length distribution from conductance probes, together with the self-organized neural network (SONN) for nonlinear mapping. Typically, neural network training requires an iterative solution with several loops [29]. In a different approach, Qi *et al.* [22] used SVM on Electrical Resistance Tomography (ERT) data for regime identification in horizontal pipes. Meanwhile, Wang and Zhang [23] used SVM on electrical capacitance tomography measurements. Trafalis *et al.* [24] applied SVM on regime data from various databases to predict flow regime transitions. In these studies, the use of kernels in SVM enables the elucidation of nonlinear boundaries between flow regimes. Thus, nonlinear SVM belongs to a class of machine learning techniques called kernel methods. Kernel methods are advantageous in terms of ease of training, simplicity of architecture, and interpretability [30].

Conductance is an appropriate intrinsic flow regime indicator [8], [16], [18], [31] since it correlates with the fraction of the liquid phase of the flow. Moreover, the instrumentation for collecting conductance data is simple, non-invasive, relatively cheap, and easy to manufacture. Yet, to our knowledge, there are currently no conductance-based objective flow regime identification tools sufficient for a real-time application in closed channels, let alone horizontal annuli. The aim of this paper is to develop such a tool.

The rest of the paper is organized as follows. We briefly describe our experimental setup in Section II. Kernel methods are detailed in Section III. Results and discussion are given in Section IV. Lastly, we conclude the work in Section V.

II. Experimental Setup

A. Two-phase Flow Loop

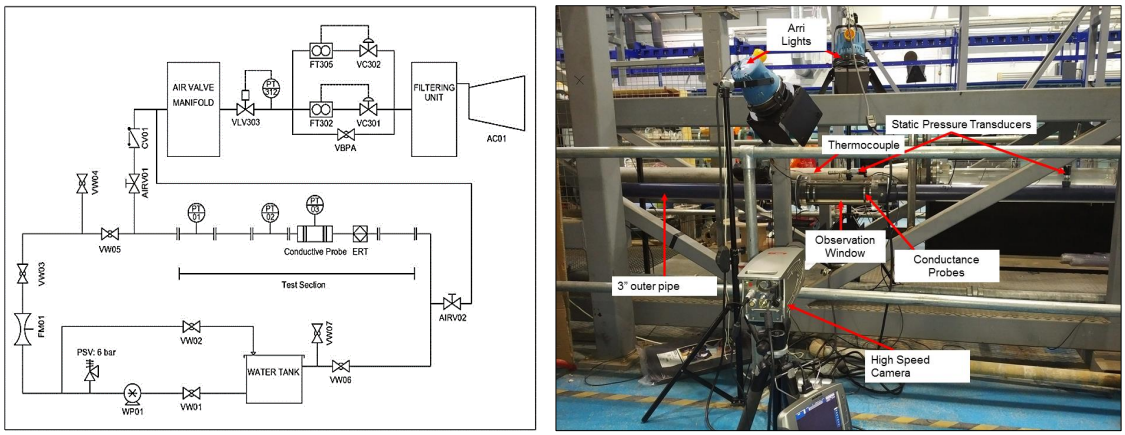
Fig. 1a shows the schematic of the experimental test rig, while Fig. 1b shows a picture of the test section. The annulus rig was designed and implemented on a rig platform in the Process Systems Engineering Laboratory at Cranfield University. The test section has an annular geometry and is made of polyvinyl chloride (PVC) pipes. The total length is 14 m, in which 2 m of 0.0765 m internal diameter is allocated for flow development before the annulus section of 10.8 m length, with 0.0768 m outer pipe diameter ($\mathrm{D_o}$) and 0.060 m ($\mathrm{D_i}$) inner pipe diameter. The inner pipe is held concentrically in the outer pipe to form the annulus using 4 mm stainless steel pins fed through an assembly designed to adjust the eccentricity of the annulus.

Air is supplied using a screw compressor by AtlasCopco®, model GA55 with a maximum discharge pressure of 7.5 barg and free air delivery maximum capacity of 638 $\mathrm{m^3/hr}$. Two automated valves (namely VC301 and VC302) are used to regulate the air flow rate, which is measured by one of two Rosemount Mass Probar flow meters (FT302 and FT305) with $\pm 1.4\%$ uncertainty. Air flow rates below 150 $\mathrm{Sm^3/hr}$ are measured by FT302, while flow rates above 150 $\mathrm{Sm^3/hr}$ are measured by FT305. Tap water is pumped from a water tank with a capacity of 2 $\mathrm{m^3}$ using a progressive cavity pump (PCP) with maximum discharge pressure of 6 barg. The water flow is measured using an Endress and Hauser Promag 50 electromagnetic flow meter ranging between 0-18 $\mathrm{m^3/hr}$, with $\pm 0.5\%$ uncertainty. Two GE Druck static pressure transducers, PMP 1400 and a WIKA model A-10 static pressure transducer ranging between 0-6 barg ($\pm 0.25\%$) of full scale are used to obtain the static pressure in the test section. The LabView® system, consisting of a National Instruments (NI) connector board interface connected to the instrumentation BNC coaxial cables, is used for data acquisition. Finally, an OLYMPUS model i-SPEED 3 high speed video camera is used to capture images at 1000 frames per second with adequate lighting supplied by Arrilite 800-W lights.

B. Conductance Probes

An electrical impedance probe consisting of two pairs of conductance rings was designed and fabricated for this study based on the recommendations of Fossa [31]. The probes are flush mounted on the outer pipe of the annulus section based on an aspect ratio of $\mathrm{D_e/D}$ of 0.34, where $\mathrm{D_e}$ is the spacing between the electrodes and D is the pipe diameter resulting in electrode spacing of 26 mm. Two pairs of electrodes were used, each having a width of 6 mm and a distance of 270 mm between each electrode pair. The probe is calibrated offline for stratified flow based on the technique employed by Fan and Yan [18]. During calibration, the uncertainty of the probe was found to be $\pm 2\%$ from direct measurements.

Voltage signals from the probe were recorded at a sampling rate of 100 Hz. Data samples were obtained over a range of gas-liquid flow rates where various flow regimes can be observed (see Section IV.A). A single data sample consists of a collection of voltage signals for 180 s, while under a certain constant setting of gas and liquid flow rate on the channel. For each data sample, a probability density function (PDF) curve, i.e. $p(v)$ vs. $V/V_{\max}$, was generated using the method in [8]. The width of each PDF bin was set to $w = 0.02$, yielding a total of $M = 50$ number of bins in a single PDF.



(a) (b)

Fig. 1. (a) Schematic of the 3-inch annulus flow loop; (b) Actual experimental setup at the test section.

III. Objective Flow Regime Identification Using Kernel Methods

Kelessidis and Dukler [8] established that features from the PDF of conductance data in the form of voltage signals can be used to distinguish among the flow regimes. However, the use of mere visual inspection of PDFs is subjective.

In this work, the first goal is to train an objective identifier from a training data set consisting of voltage signal PDF data at selected gas-liquid flow rate settings, together with their assigned flow regime labels. During training, a mapping can be learned between the voltage signal PDFs and their flow regime labels. When learned, the patterns from the training data can be generalized to any gas-liquid flow rate setting, thus providing an objective flow regime identification tool. After which, the final goal is to assess the objective identifier for real-time application.

In this section, we detail our approach of using kernel methods for achieving the first goal. The objective identifier itself consists of a kernel principal components analysis (KPCA) based mapping of every sample PDF curve in some 2-dimensional feature space, followed by a kernel support vector machine (SVM) classifier for computing transition boundaries between flow regimes in the 2D feature space.

A. Kernel Principal Components Analysis

In this work, each bin of $V/V_{\max}$ in a PDF curve is treated as a dimension or "feature". For a PDF curve with M bins, a dimensionality reduction method is used to project M features to 2 features, which are then made to comprise the coordinates of some 2-dimensional plane. This step is *unsupervised*, i.e. it proceeds without the knowledge of the flow regime labels yet. After this step, every PDF derived from each data sample is now represented as a data point on a 2D map, which we call the *virtual flow regime map* throughout this paper.

The virtual flow regime map must retain the properties of actual flow regime maps in that: (i) when the data points are given labels, a clear separation should emerge between the points from different flow regimes; and, (ii) when gas and liquid flow rates are steadily increased or decreased during real-time use, a continuous path of points must be traceable in the map by successively projecting incoming PDF curves from a moving time-window of voltage signals.

Principal components analysis (PCA) is a standard linear dimensionality reduction tool that aims to project input data onto a lower dimensional *latent* space, while preserving information in the form of variance of data. For instance, Wang and Zhang [23] have used PCA for mapping electrical capacitance data prior to flow regime classification. Going further, kernel PCA is a nonlinear extension of PCA, proposed by Scholkopf *et al.* [32], that enhances the separation of dissimilar data in the latent space. In this work, the following KPCA methodology from [32] is adopted.

Let $\mathbf{x}_i \in \Re^M, i = 1, 2, \ldots, N$ denote any training data set of N samples of M features. Initially, all features are normalized to zero mean and unit variance. KPCA then proceeds by forming an $N \times N$ kernel matrix,

$$\mathbf{K} = [K_{ij}] = [K(\mathbf{x}_i, \mathbf{x}_j)] \tag{1}$$

using a suitable kernel function $K(\cdot, \cdot)$, which roughly acts as a *similarity measure* between any pair of samples, $\mathbf{x}_i$ and $\mathbf{x}_j$. In this work, the widely used radial basis function (RBF) kernel is adopted, given by

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{k_w}\right) \tag{2}$$

where k_w is the kernel width. Note that the value of K tends to 1.0 when $\mathbf{x}$ and $\mathbf{x}'$ are similar, otherwise it tends to zero. The kernel matrix, K is then centered as,

$$\mathbf{K}_c = \mathbf{K} - \mathbf{1}_N\mathbf{K} - \mathbf{K}\mathbf{1}_N + \mathbf{1}_N\mathbf{K}\mathbf{1}_N \tag{3}$$

where $\mathbf{1}_N$ is an $N \times N$ matrix and $(\mathbf{1}_N)_{ij} = 1/N$.

The eigenvalue decomposition of $\mathbf{K}_c$ is

$$\mathbf{K}_c/N = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \tag{4}$$

where each column of $\mathbf{V} \in \Re^{N\times N}$ is an eigenvector and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots \lambda_N) \in \Re^{N\times N}$ is a diagonal matrix of

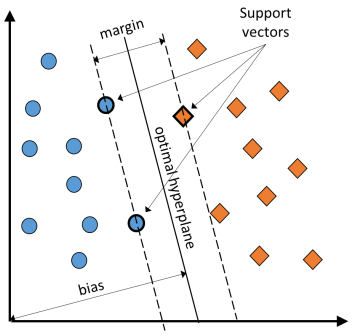


Fig. 2. Illustration of optimal separating hyperplane in SVM.

decreasing eigenvalues. From (4), the projection matrix is formed as

$$\mathbf{P} = \mathbf{V}\mathbf{\Lambda}^{-1/2} \in \Re^{N\times N} \tag{5}$$

where $\mathbf{\Lambda}^{-1/2}$ is used to scale the eigenvectors to ensure orthogonality of the kernel principal components [33].

By extracting the first 2 columns of $\mathbf{P}$, denoted as $\mathbf{P}_2$, only the top 2 projections that capture the maximum variance in the data are effectively retained. These projections, also known as *latent variables* $\mathbf{L}$, are the new features in the latent space:

$$\mathbf{L} = \mathbf{P}_2^T\mathbf{K}_c \in \Re^{2\times N}. \tag{6}$$

Together with their respective labels $y_i, i = 1, 2, \ldots, N$ from the set of observed flow regimes, the $\mathbf{L}$ becomes input to SVM for subsequent flow regime classification.

In this work, only 2 features are extracted since: (i) data visualization is desired; (ii) overfitting may occur when too many features are chosen but N is small; and, (iii) it is known that only 2 variables are being changed to generate all the data, namely, the gas and liquid flow rates. In other words, the *intrinsic* dimensionality of the data is known to be 2. Thus, 2 features is enough to retain flow-distinguishing information from PDF curves. Overfitting is also prone to occur if all M features of the PDF are sent to SVM without the dimensionality reduction step, e.g. PCA or KPCA. If SVM overfits, the identifier may perform too well on the training data, but poorly for unseen test data.

B. Binary Kernel Support Vector Machine

After visualizing all PDF data in 2D space, data points belonging to the same flow regimes are expected to cluster together. SVM can then be used to establish exact boundaries between different flow regime clusters. The benefits of SVM are two-fold: (i) the method itself is based on the Structural Risk Minimization (SRM) principle, as opposed to Empirical Risk Minimization (ERM) employed by classical least-squares techniques, thus, SVM is a flexible learner and has good generalizability [29], [30]; and (ii) training an SVM only involves solving a convex quadratic programming problem, which is known to have a unique solution. In contrast, ANNs are usually trained iteratively by gradient descent, wherein the solution might converge to local minima [29].

Originally, Cortes and Vapnik [34] proposed the SVM for classifying data into 2 classes (binary), which are labelled as $y \in \{+1, -1\}$. The SRM principle is employed by searching for a linear separating hyperplane that maximizes the margin between the positive and negative classes, as illustrated in Fig. 2. Nonlinear boundaries are enabled by using kernels in a dual form of the SVM [30]. In this paper, the so-called 1-norm C-SVM is adopted, which is posed as the following convex quadratic programming dual problem [30]:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to} \quad & \sum_{i=1}^{N} y_i \alpha_i = 0 \\ & 0 \le \alpha_i \le C \quad i = 1, 2, \ldots, N \end{aligned} \tag{7}$$

where $\mathbf{x}_i$ are the N training samples (each column of $\mathbf{L}$ in (6)), α_i are dual variables, $K(\cdot, \cdot)$ is some kernel function, $y_i \in \{+1, -1\}$ are the known labels, and C is a regularization constant called the *box constraint*. The theory and derivation of this form of SVM is found in [30].

Solving (7) gives the optimal values α_i^*, from which the decision function is obtained as

$$f(\mathbf{x}) = \sum_{i\in SV} y_i \alpha_i^* K(\mathbf{x}_i, \mathbf{x}) + b^* \tag{8}$$

where b^* is a bias term and SV is a subset of training samples that emerged as *support vectors*, i.e. the samples that define decision boundaries just as the ones highlighted in Fig. 2. The bias, b^*, is chosen so that $y_i f(\mathbf{x}_i) = 1$ for any i with $0 < \alpha_i^* < C$ [30]. By obtaining $f(\mathbf{x})$, the SVM classifier is now said to be *trained*. Hence, for any new sample $\mathbf{x}$, the decision rule is given by $y = \text{sign}(f(\mathbf{x}))$ which returns $+1$, -1, or 0. The samples, $\mathbf{x}$, for which $\text{sign}(f(\mathbf{x})) = 0$ constitutes the exact boundary between the two classes.

In this work, the RBF kernel (2) is also used as $K(\cdot, \cdot)$ in (7), but it can have a different k_w than the one used in KPCA. We refer to k_w and C in SVM as *hyper-parameters* in this paper. Since k_w and C are problem-dependent, their values must be defined on a case-to-case basis. Typically, a range of values of k_w and C is initially investigated to search for the best SVM classification performance [30]. Overfitting may occur when k_w is too low, while decision boundaries tend to be more linear at high k_w, at the expense of accuracy. Similarly, the value of C controls regularization: When $C = \infty$, a strict separation between classes is enforced (hard margin SVM). But as a result, decision boundaries become more complex, sometimes leading to non-contiguous regime regions. Otherwise for $0 < C < \infty$, the boundaries tend to be simpler, since α_i is being restricted within a range. But as C is decreased, more and more misclassified samples are being tolerated (soft margin SVM).

C. Multi-class Kernel Support Vector Machine

Since there are more than 2 flow regimes, multi-class SVMs must be employed. This involves training many binary SVM classifiers according to either one-versus-one (1-v-1), one-versus-rest (1-v-r), or max-wins strategies. In this work, we

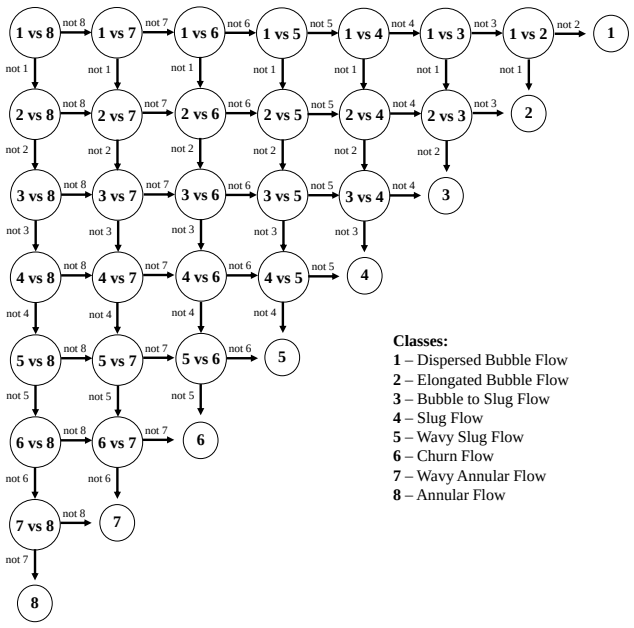


Fig. 3. DDAG of binary SVM classifiers for flow regime classification in the experimental setup of this study.

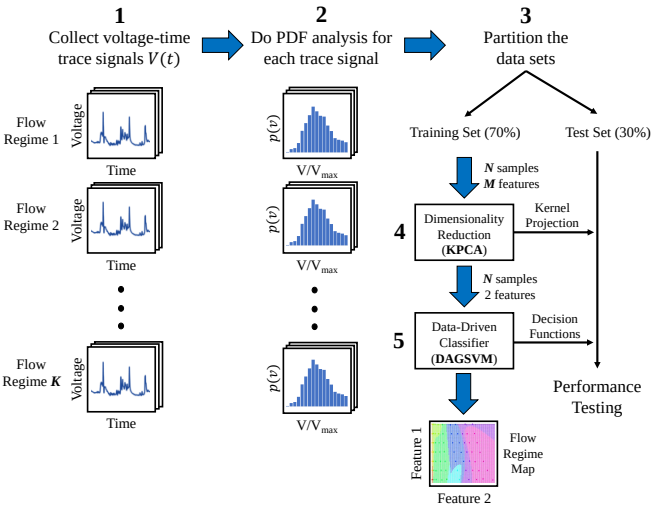


Fig. 4. Proposed objective flow regime identification approach using conductance data and kernel methods.

adopt an efficient 1-v-1 algorithm called DAGSVM, proposed by Platt *et al.* [35].

In DAGSVM, $S(S-1)/2$ binary SVM classifiers are trained between every possible pair of distinct classes, where S is the number of classes or flow regimes. In each classifier, the two competing classes are assigned the $+1$ and -1 labels, and then a decision boundary is trained between them. In this paper, the k_w and C values for all classifiers are set to be the same, since they are equally important. The resulting $S(S-1)/2$ decision functions are represented in each node of a decision directed acyclic graph (DDAG), shown in Fig. 3 for $S = 8$.

To use DAGSVM, any incoming sample is classified first in the root node, 1 vs. 8. From the current node, the DDAG is traversed either downwards or to the right. The sample is further classified successively on every node in the path, until a final decision is reached on a leaf node. This approach is efficient since any path from the root to any leaf has a length of $S-1$. Thus, a sample needs to be evaluated only $S-1$ times before it is conclusively classified to a flow regime.

In summary, the methodology for objective flow regime identification using conductance data and kernel methods is outlined in Fig. 4. In the end, a virtual flow regime map is sought from the data, where transitions between flow regimes are depicted by exact boundaries. However, "virtual" means that the axes of this map have no physical meaning and the computed exact boundaries by the SVM have no physical basis as well. Rather, principles from machine learning are employed in place of a physical basis. Hence, our approach is mainly data-driven. Nevertheless, the virtual flow regime map is a means for our objective flow regime identifier to visually explain its results. This interpretability feature is the advantage of our approach over existing data-driven approaches [14].

IV. Results and Discussion

In this section, we first present results for the static performance of KPCA-SVM based objective flow regime identifiers. Results for real-time identification follow thereafter.

A. Assignment of flow regime labels

A total of 110 data samples were obtained from the experimental setup (see Section II) at various air and water flow rates within 0.17-22.8 m/s and 0.14-2.8 m/s, respectively. Flow regime labels were assigned to each sample by a human expert at the same ambient conditions, using a frame-by-frame visual analysis of video recordings from the high-speed, high-fidelity camera. Regime transitions were also determined after consulting other experts. Eight distinct flow regimes were observed, namely: Dispersed Bubble Flow, Elongated Bubble Flow, Bubble to Slug Flow, Slug Flow, Wavy Slug Flow, Churn Flow, Wavy Annular Flow, and Annular Flow. Due to limitations in the experimental setup, the Stratified Flow regime was not encountered. Nonetheless, this regime may be encountered at lower gas-liquid flow rates adjacent to the Annular Flow region in a different setup. If encountered, the DDAG in Fig. 3 must be extended to 9 classes.

B. From voltage signals to PDF curves

Fig. 5 shows typical PDF curves analysed from the voltage signals training data from each flow regime. It can be seen that the PDF curves generally skew from high voltage values to low voltage values as one moves from Dispersed Bubble Flow to Annular Flow in order. Compared to previous works, similar trends were observed at the ranges of gas and liquid flow rates considered in annular channels [8].

The data set of PDF curves and their respective flow regime labels was then partitioned into $N = 77$ samples for training, and 33 samples for testing. In the subsequent KPCA of the training data, the aforementioned patterns in the PDF curves is to be extracted so that the arrangement of data points in 2D KPCA latent space can reveal directions where gas and liquid flow rates increase or decrease.

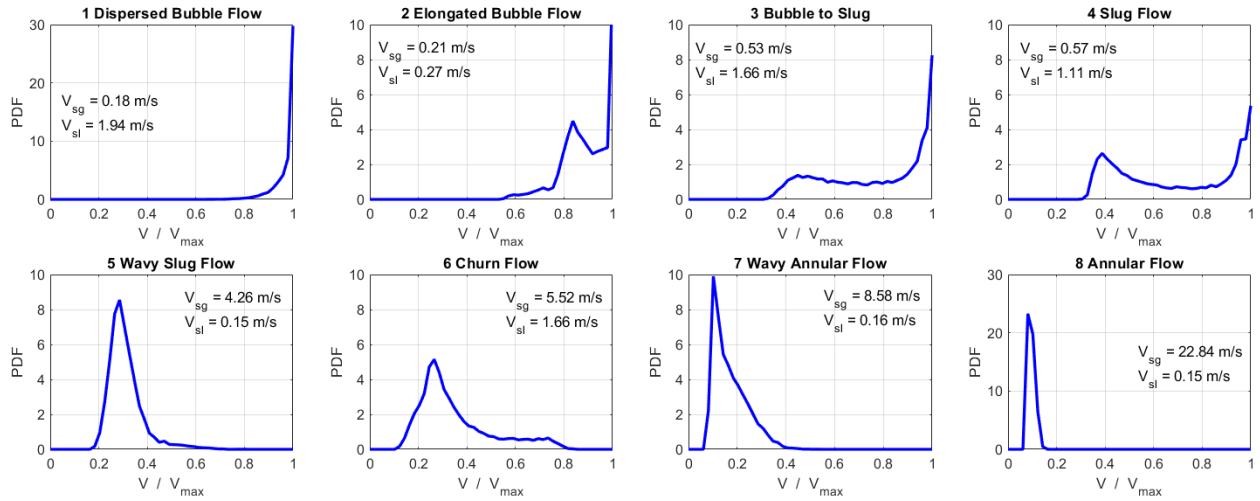


Fig. 5. Typical PDF curves from each flow regime.

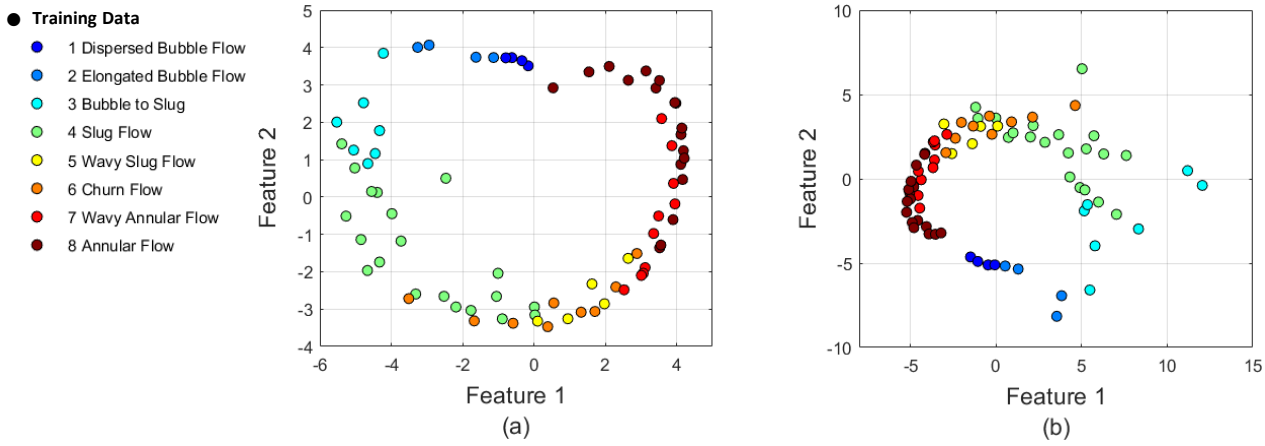


Fig. 6. (a) KPCA and (b) PCA 2D visualization of probability density function (PDF) features from the conductance training data.

C. From PDF curves to latent space

From the PDF curves, a 2-dimensional map in the latent space was produced by applying KPCA on the training samples, shown in Fig. 6a. The kernel width k_w was chosen as 0.8 quantile of the $||\mathbf{x}_i - \mathbf{x}_j||^2$ statistics [36]. Linear PCA results are also shown in Fig. 6b for comparison. The advantage of KPCA over PCA is evident in that samples from different flow regimes are more clearly separated in KPCA.

Somehow, an intuition can be built for the resulting KPCA latent space in Fig. 6a. First note that none of the points are located near the origin, $(0, 0)$. A test sample near the origin would correspond to a PDF curve where the $p(v)$ features lie *close* to their respective means. The training samples are found radially around the map since their PDF features deviate from the mean values in a specific way. For example, Slug, Bubble, and Annular Flow regime samples emerged on the lower-left, upper-left, and upper-right regions of the map, respectively. Hence, it is possible to reveal the various flow regime clusters in the 2D KPCA visualization of conductance data.

Ambiguity between some Churn flow, Wavy Slug Flow, and Slug Flow samples occurs in the lower-right region of the map. The same ambiguity can be visually observed in the PDF curves of these samples (see Fig. 5), in that they all possess a dominant peak at the region $0.2 \leq V/V_{\max} \leq 0.4$. Due to this, soft margin SVM classifiers were used to allow for misclassification.

D. From latent space to virtual flow regime map

Multi-class DAGSVM was then used to establish the boundaries objectively between flow regimes in the latent space. The MATLAB code used in this work is available in [37].

A sample flow regime map for $k_w = 3$ and $C = 100$ is shown in Fig. 7a, where training and test samples are denoted as circles and triangles, respectively. Basically, SVM assigned every location in the latent space to a flow regime by measuring its similarity to all neighboring training data locations where the flow regimes are known. As explained in Section III.A, this *similarity measure* is exactly what the kernel in (2) computes, when called by the decision function in (8). As a result, maximally separating nonlinear boundaries are produced. This is the logic offered by our approach in place of human subjectivity in flow regime classification. Arguably, results can improve with more training data samples.

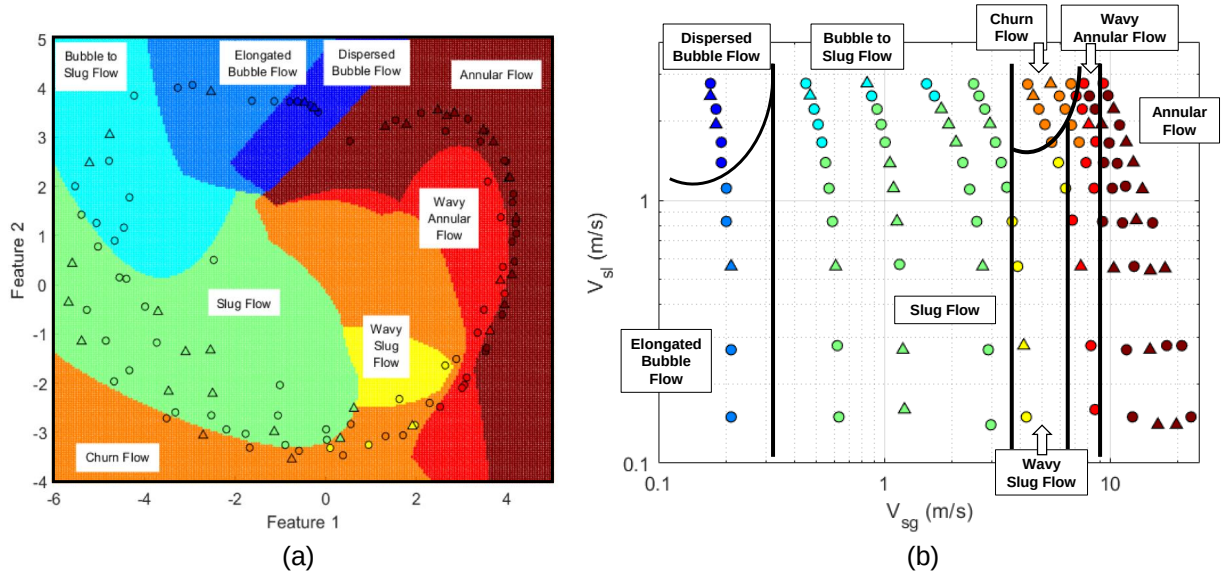


Fig. 7. (a) Virtual flow regime map from conductance data using SVM ($k_w = 3$; $C = 100$); (b) Actual flow regime map from gas-liquid flow rates data with arbitrarily drawn flow regime boundaries. Training samples are denoted by circles, while test samples are denoted by triangles. For (a), if the color of a sample is mismatched with the color of the region where it lies, the sample is misclassified.

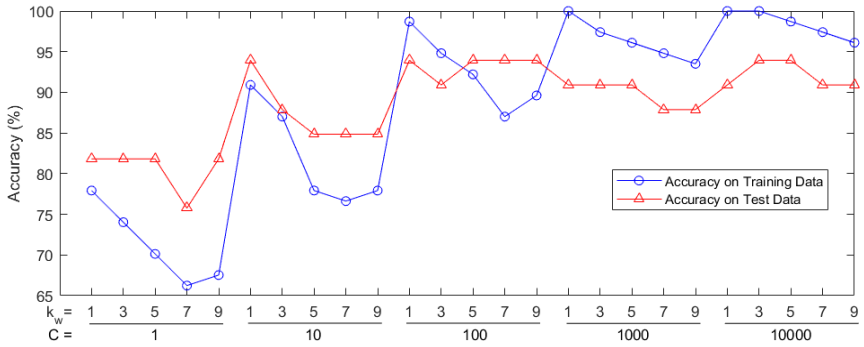


Fig. 8. Accuracy of SVM classifiers at various k_w and C settings.

At the hyper-parameter settings in Fig. 7a, the classification accuracy against visually observed flow regimes for training and test data are 94.8% and 90.9%, respectively. In comparison, Fig. 7b shows a typical flow regime map from the gas-liquid flow rates data where the flow regime boundaries are drawn arbitrarily. These results demonstrate the effectiveness of our approach for objective flow regime classification.

Misclassified data samples in Fig. 7a are listed in Table I. The objective identifier encountered difficulties in distinguishing between some Wavy Slug Flow and Churn Flow samples. This can be explained by the resemblance found between the PDF curves of some Wavy Slug Flow and Churn Flow samples. However, a close inspection of the acquired voltage signals actually reveals distinct patterns between these two flow regimes. Hence, this misclassification problem can be attributed to the fact that PDF curves simply account for the static locations of voltage values. The identifiability of various intermittent flow conditions can then be improved by adding features related to the *dynamics* in the voltage time series. With this, a clearer separation can be expected among different intermittent flow conditions in the latent space.

TABLE I
MISCLASSIFIED SAMPLES IN THE FLOW REGIME MAP IN FIG. 7A.

A. Misclassified training samples (4 out of 77)			
V_{sg} (m/s)	V_{sl} (m/s)	SVM Decision	Expert's Decision
3.69	0.83	Churn Flow	Wavy Slug Flow
3.90	0.56	Churn Flow	Wavy Slug Flow
4.26	0.15	Churn Flow	Wavy Slug Flow
8.65	1.67	Annular Flow	Wavy Annular Flow
B. Misclassified test samples (3 out of 33)			
V_{sg} (m/s)	V_{sl} (m/s)	SVM Decision	Expert's Decision
1.23	0.16	Churn Flow	Slug Flow
4.15	0.28	Churn Flow	Wavy Slug Flow
8.05	1.94	Annular Flow	Wavy Annular Flow

Misclassification also raises the issue of overfitting vs. underfitting, which depends on the choice of SVM hyper-parameters. As mentioned, different settings for k_w and C produce varying classification performance. The accuracy of

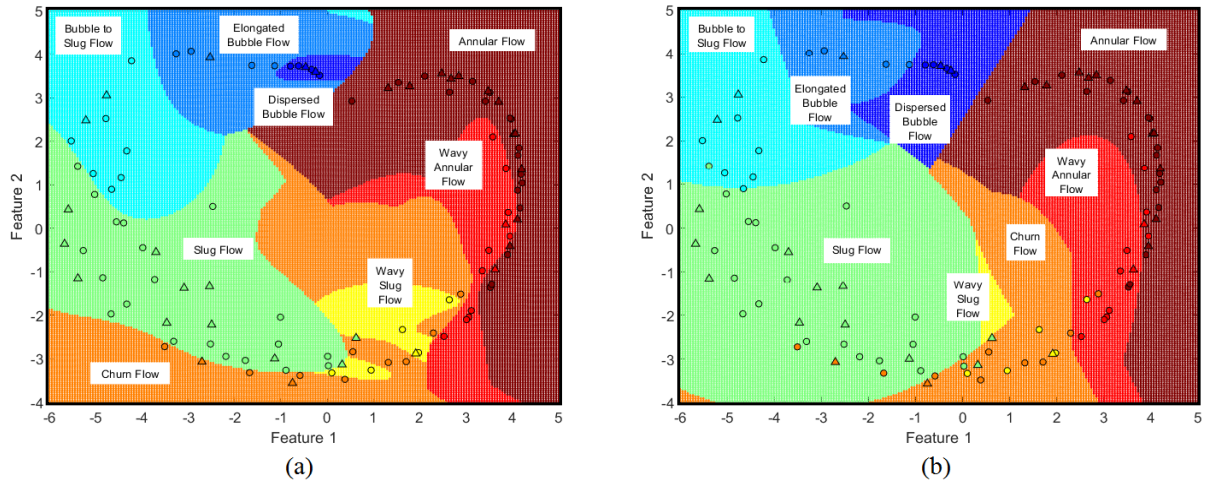


Fig. 9. Virtual flow regime maps from conductance data using SVM: (a) overfit ($k_w = 1$, $C = 10^4$); and (b) underfit ($k_w = 9$, $C = 10$). Training samples are denoted by circles, while test samples are denoted by triangles. If the color of a sample is mismatched with the color of the region where it lies, the sample is misclassified.

SVM over a grid of hyper-parameter values, namely $k_w \in \{1, 3, \ldots, 9\}$ and $C \in \{1, 10, \ldots, 10^4\}$, is depicted in Fig. 8. By adjusting k_w and C towards overfitting (lower k_w and higher C), the classifier becomes biased towards the human expert decisions, giving a high classification accuracy on the training data but a low accuracy on test data. One case of overfitting is illustrated by the flow regime map in Fig. 9a for $k_w = 1$ and $C = 10^4$. Overfitting is evidenced by tighter and more complex boundaries, indicating that the classifier tends to memorize, rather than generalize, the patterns from the training data samples. On the other hand, in a particular case of high k_w and low C in Fig. 9b, underfitting occurs. In the case of underfitting, the created boundaries tend to be too simple, leading to low classification accuracies in both training and test data.

After evaluating the accuracy and visually checking the sensibility of the virtual flow regime maps for each combination, $k_w = 3$ and $C = 100$ were finally chosen as the values that balance overfitting and underfitting. Thus, these settings were used in the objective classifier for real-time application. The task of avoiding over- and underfitting via hyper-parameter tuning, i.e. controlling the model complexity, is common to all supervised learning machines, including ANNs. In the future, larger training data sets must be generated for a more thorough validation of classification performance.

E. Application to real-time flow regime identification

The final goal of this paper is to implement the trained KPCA-SVM identifier for real-time automatic flow regime identification.

In a new experiment, the liquid flow rate going through the annular pipe was fixed at 0.95 m/s while the gas flow rate was being increased steadily from 1.48 m/s to 26.17 m/s. The trajectory of gas-liquid flow rate data is plotted in Fig. 10c, where a blue circle marks every 5 min. The positions of the training samples are also superimposed in Fig. 10c. The total duration of the experiment is 30 min. To contrast with the *virtual* flow regime map in Fig. 10b, we will refer to Fig. 10c as the *actual* flow regime map. Throughout the real-time experiment, voltage signals were being collected at 100 Hz sampling rate, as presented in Fig. 10a.

Real-time identification involves the continuous processing of a moving window of voltage signals. As noted in Sec. II.B, the length of this window is set as 180 s. By moving the window every second, the latest 180-second voltage signal data set is turned into a single PDF curve to be processed by the KPCA-SVM identifier. Hence, a flow regime is identified at the end of every second as well. Note that one needs to wait for the first 180 seconds to elapse in order to form the first window. To summarize all identified flow regimes, a *regime chart* is introduced in Fig. 10a. The y-axis of the chart lists all possible flow regime classes, while the x-axis is time. This chart automatically plots the predominant flow regime in the flow channel against time.

The effectiveness of our method is demonstrated by observing the trajectory of the experiment in Fig. 10c and the results in Fig. 10a. With only the knowledge of the flow regimes provided by the training data, our method was able to infer the flow regime at different gas-liquid flow rates both within and beyond the vicinity of the training data. For example, the first detected regime was Slug Flow, and this result agrees with the fact that the experimental run started with a data point close to where the Slug Flow samples are located in the actual flow regime map (Fig. 10c). As one moves in the trajectory of data points, various flow regimes must have been crossed, and these transitions were reflected in the regime chart smoothly across time. Thus, an identifier which reads conductance data and determines the flow regime objectively in real-time has been developed, as opposed to previous subjective means.

The trajectory of the data points is also plotted in the virtual flow regime map given in Fig. 10b. This is the same map presented Fig. 7a, but depicting only the boundaries between flow regimes. The main benefit of the virtual flow regime map is that it gives an explanation why the objective

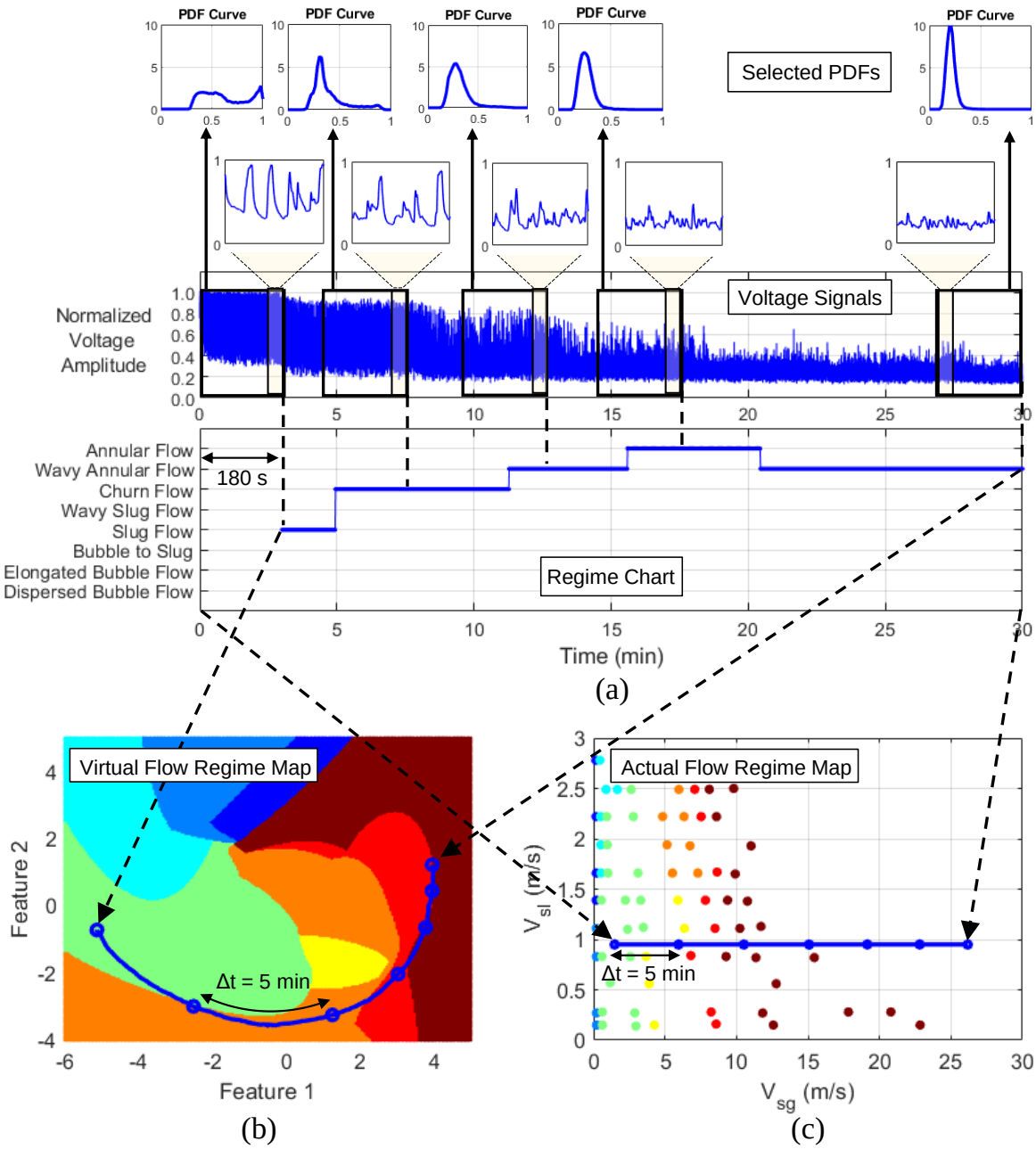


Fig. 10. Results for real-time flow regime identification: (a) Selected PDF curves, actual voltage signals, and regime chart; (b) trajectory of data points in the virtual flow regime map; (c) trajectory of data points in the actual flow regime map, along with training samples. Colors are adopted from Fig. 7a.

identifier predicted a certain flow regime. For example, at 5 min, our identifier changed its decision from Slug Flow to Churn Flow. This is because in the virtual flow regime map, the trajectory has just crossed the exact boundary between Slug Flow and Churn Flow calculated by the SVM. Due to these results, the constructed virtual flow regime map via KPCA is seen to be a useful visualization of conductance data. Hence, our approach overcomes the disadvantage of other black-box pattern recognition models, such as ANNs, that give no visual aid as to how they arrived at their decisions.

F. Final remarks and future implications

Some remarks about our objective identifier are given as follows. Note that a time-window length of 180 s was used for creating PDF curves from voltage signals (see Section II.B). Indeed, this is an arbitrary choice, yet, it has a high impact to regime identification performance. Too large a window length may delay the response time of identification, while too low admits noise. However, the window length can also be viewed as a tuning parameter, which can be called the *resolution*. Two objective identifiers can be trained: one having a higher resolution (smaller window length) than the other. When used together in real-time applications, the discrepancies between

the decisions of each identifier can reveal more information about the multiphase flow. For example, one might be interested in detecting bubbles using a high-resolution identifier, while a low-resolution identifier simultaneously reads Annular Flow. The width of the PDF bins, w, could also be a parameter for adjusting resolution.

Another potential application of this work is to enable switching hybrid models for multiphase flow. One can use the regime chart results to switch between different flow-regime-specific models for estimating flow properties or for model-based control. Low-resolution identifiers can be used to avoid too many switches.

This work can also be improved in several ways. For example, more attention must be given to the feature extraction and dimensionality reduction steps of our method. In the future, it is worthwhile to find feature extraction methods, other than the use of PDF features, that can better distinguish among the flow regimes, especially intermittent flow. Also, to improve performance tuning, it is envisaged that further experimental work will provide additional data sets for training, validation and testing in each of the 8 flow regime domains. Lastly, further tests are needed to assess the repeatability and scalability of our method to other systems. If system-specific issues arise, the feature extraction step is most likely to adapt since KPCA and SVM are the steps in the overall approach which are portable to any scenario. Hence, many other research activities can originate from this work.

V. Conclusion

In this paper, an objective flow regime identifier for gas-liquid two-phase flow in horizontal annuli was developed using kernel methods, namely, kernel principal component analysis (KPCA) and multi-class kernel support vector machine (SVM). The objective identifier solely uses conductance data and its PDF features for classifying flow regimes, without the knowledge of the gas and liquid flow rates through the channel. At chosen SVM hyper-parameters that avoid overfitting and underfitting, the accuracy of our objective identifier against visually observed flow regimes for training and test samples were 94.8% and 90.9%, respectively.

Furthermore, our method has enabled an efficient real-time automatic flow regime identification tool using a moving window of solely conductance data for the first time. We have introduced the concept of *regime charts* for reporting the identified predominant flow regime every second, as well as *virtual flow regime maps* for tracking how the identifier arrived to its decisions in real-time. Our work has many potential implications in the area of multiphase flow. Nonetheless, there is much room for improvement in our method, such as the use of more advanced feature extraction and dimensionality reduction techniques that can better distinguish samples from different flow regimes.

References

[1] B. Wu, M. Firouzi, T. Mitchell, T. E. Rufford, C. Leonardi, and B. Towler, "A critical review of flow maps for gas-liquid flows in vertical pipes and annuli," *Chem. Eng. J.*, vol. 326, pp. 350–377, 2017.

[2] L. Liebenberg and J. P. Meyer, "Objective classification of two-phase flow regimes," *Heat Transf. Eng.*, vol. 29, no. 1, pp. 1–2, 2008.

[3] C. Shanthi, N. Pappa, and J. A. Suganya, "Digital Image Processing Based Flow Regime Identification of Gas/Liquid Two - Phase Flow," *IFAC Proc. Vol.*, vol. 46, no. 32, pp. 409–414, dec 2013.

[4] Y. Zhou, F. Chen, and B. Sun, "Identification Method of Gas-Liquid Two-phase Flow Regime Based on Image Multi-feature Fusion and Support Vector Machine," *Chinese J. Chem. Eng.*, vol. 16, no. 6, pp. 832–840, 2008.

[5] G. Roshani, E. Nazemi, S. Feghhi, and S. Setayeshi, "Flow regime identification and void fraction prediction in two-phase flows based on gamma ray attenuation," *Measurement*, vol. 62, pp. 25–32, feb 2015.

[6] F. Dong, X. Liu, X. Deng, L. Xu, and L.-a. Xu, "Identification of two-phase flow regimes in horizontal, inclined and vertical pipes," *Meas. Sci. Technol.*, vol. 12, no. 8, pp. 1069–1075, 2001.

[7] J. E. Julia, B. Ozar, J. J. Jeong, T. Hibiki, and M. Ishii, "Flow regime development analysis in adiabatic upward two-phase flow in a vertical annulus," *Int. J. Heat Fluid Flow*, vol. 32, no. 1, pp. 164–175, 2011.

[8] V. C. Kelessidis and A. E. Dukler, "Modeling flow pattern transitions for upward gas-liquid flow in vertical concentric and eccentric annuli," *Int. J. Multiph. Flow*, vol. 15, no. 2, pp. 173–191, 1989.

[9] G. Das, P. Das, N. Purohit, and A. Mitra, "Flow Pattern Transition During Gas Liquid Upflow Through Vertical Concentric Annuli Part II: Mechanistic Models," *J. Fluids Eng.*, vol. 121, no. 4, pp. 902–907, 1999.

[10] R. E. Osgouei and E. Ozbayoglu, Evren M, Ozbayoglu, Murat A., Yuksel, "Flow Pattern Identification of Gas-Liquid Flow Through Horizontal Annular Geometries," *Onepetro*, 2010.

[11] S. Wongwises and M. Pipathattakul, "Flow pattern, pressure drop and void fraction of two-phase gas-liquid flow in an inclined narrow annular channel," *Exp. Therm. Fluid Sci.*, vol. 30, no. 4, pp. 345–354, 2006.

[12] N. Ekberg, S. Ghiaasiaan, S. Abdel-Khalik, M. Yoda, and S. Jeter, "Gasliquid two-phase flow in narrow horizontal annuli," *Nucl. Eng. Des.*, vol. 192, no. 1, pp. 59–80, aug 1999.

[13] E. N. Eyo and L. Lao, "Gas-liquid flow regimes in horizontal annulus," *J. Pet. Sci. Eng.*, vol. 175, pp. 573–586, 2019.

[14] Y. Yan, L. Wang, T. Wang, X. Wang, Y. Hu, and Q. Duan, "Application of soft computing techniques to multiphase flow measurement: A review," *Flow Meas. Instrum.*, vol. 60, pp. 30–43, 2018.

[15] Y. Mi, M. Ishii, and L. H. Tsoukalas, "Flow regime identification methodology with neural networks and two-phase flow models," *Nucl. Eng. Des.*, vol. 204, no. 1-3, pp. 87–100, 2001.

[16] L. Hernández, J. E. Juliá, S. Chiva, S. Paranjape, and M. Ishii, "Fast classification of two-phase flow regimes based on conductivity signals and artificial neural networks," *Meas. Sci. Technol.*, vol. 17, no. 6, pp. 1511–1521, 2006.

[17] C. M. Salgado, C. M. Pereira, R. Schirru, and L. E. Brandão, "Flow regime identification and volume fraction prediction in multiphase flows by means of gamma-ray attenuation and artificial neural networks," *Prog. Nucl. Energy*, vol. 52, no. 6, pp. 555–562, 2010.

[18] S. Fan and T. Yan, "Two-Phase Air-Water Slug Flow Measurement in Horizontal Pipe Using Conductance Probes and Neural Network," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 2, pp. 456–466, feb 2014.

[19] G. H. Roshani, E. Nazemi, and M. M. Roshani, "Identification of flow regime and estimation of volume fraction independent of liquid phase density in gas-liquid two-phase flow," *Prog. Nucl. Energy*, vol. 98, pp. 29–37, 2017.

[20] G. Roshani, E. Nazemi, and M. Roshani, "A novel method for flow pattern identification in unstable operational conditions using gamma ray and radial basis function," *Appl. Radiat. Isot.*, vol. 123, pp. 60–68, may 2017.

[21] G. H. Roshani, R. Hanus, A. Khazaei, M. Zych, E. Nazemi, and V. Mosorov, "Density and velocity determination for single-phase flow based on radiotracer technique and neural networks," *Flow Meas. Instrum.*, vol. 61, pp. 9–14, 2018.

[22] G.-H. Qi, F. Dong, Y.-B. Xu, M.-M. Wu, and J. Hu, "Gas/Liquid Two-phase Flow Regime Identification in Horizontal Pipe Using Support Vector Machines," in *4th Int. Conf. Mach. Learn. Cybern.*, Guangzhou, 2005.

[23] H. X. Wang and L. F. Zhang, "Identification of two-phase flow regimes based on support vector machine and electrical capacitance tomography," *Meas. Sci. Technol.*, vol. 20, no. 11, 2009.

[24] T. B. Trafalis, O. Oladunni, and D. V. Papavassiliou, "Two-Phase Flow Regime Identification with a Multiclassification Support Vector Machine (SVM) Model," *Ind. Eng. Chem. Res.*, vol. 44, no. 12, pp. 4414–4426, jun 2005.

[25] C. Alimonti and G. Falcone, "Integration of Multiphase Flowmetering, Neural Networks, and Fuzzy Logic in Field Performance Monitoring," *SPE Prod. Facil.*, vol. 19, no. 01, pp. 25–32, feb 2004.

[26] D. Xie, Z. Huang, H. Ji, and H. Li, "An Online Flow Pattern Identification System for GasOil Two-Phase Flow Using Electrical Capacitance Tomography," *IEEE Trans. Instrum. Meas.*, vol. 55, no. 5, pp. 1833–1838, oct 2006.

[27] R. Banasiak, R. Wajman, T. Jaworski, P. Fiderek, H. Fidos, J. Nowakowski, and D. Sankowski, "Study on two-phase flow regime visualization and identification using 3D electrical capacitance tomography and fuzzy-logic classification," *Int. J. Multiph. Flow*, vol. 58, pp. 1–14, 2014.

[28] B. M. Abbagoni and H. Yeung, "Non-invasive classification of gas-liquid two-phase horizontal flow regimes using an ultrasonic Doppler sensor and a neural network," *Meas. Sci. Technol.*, vol. 27, no. 8, 2016.

[29] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2008.

[30] N. Cristianini and J. Shawe-Taylor, *Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2014.

[31] M. Fossa, "Design and performance of a conductance probe for measuring the liquid fraction in two-phase gas-liquid flows," *Flow Meas. Instrum.*, vol. 9, no. 2, pp. 103–109, jun 1998.

[32] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

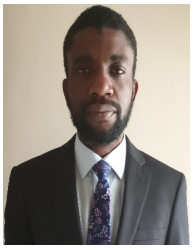
[33] S. W. Choi, C. Lee, J. M. Lee, J. H. Park, and I. B. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," *Chemom. Intell. Lab. Syst.*, vol. 75, no. 1, pp. 55–67, 2005.

[34] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[35] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification," *Int. Conf. Neural Inf. Process. Syst.*, pp. 547–553, 2000.

[36] A. Karatzoglou, D. Meyer, and K. Hornik, "Support Vector Machines in R," *J. Stat. Softw.*, vol. 15, no. 9, pp. 1–28, 2006.

[37] K. Pilario, "Binary and Multi-class SVM," 2018. [Online]. Available: https://uk.mathworks.com/matlabcentral/fileexchange/65232-binary-and-multi-class-svm



Edem Nsefik Eyo received the B.Eng. degree in Chemical Engineering from the University of Uyo, Nigeria in 2009. He also holds an M.Sc degree in Petroleum and Gas Engineering (with Distinction) from the University of Salford, Manchester, U.K., in 2013, and a Ph.D. in Energy from Cranfield University, Cranfield, U.K., in 2019.

He joined the Department of Mechanical engineering at the Cross River University of Technology as a Lecturer in 2011 and was a part of the Energy research group. His research interests are in flow assurance, multiphase flow systems and experimental fluid dynamics. His expertise is in multiphase flow in complex geometries.

Karl Ezra S. Pilario received the B.Sc., *Summa Cum Laude*, and M.Sc. degrees both in Chemical Engineering from the University of the Philippines, Diliman, in 2012 and 2015, respectively. He is currently pursuing a Ph.D. in Energy at Cranfield University, Cranfield, U.K.

He is an Assistant Professor on study leave from the Department of Chemical Engineering at the University of the Philippines, Diliman. In 2015, he received the Limcaoco Young Instructor Award for teaching excellence, undergraduate advising, and mentoring in academic competitions. He also held the Dominador I. Ilio Engineering Centennial Professorial Chair from 2015 to 2017. His research interests are in industrial applications of data analytics and machine learning, including Process Monitoring.

Liyun Lao received the B.Sc. and M.Sc. degrees in the measurement science and technology, and Ph.D. degree in the process and control engineering from Zhejiang University, China, in 1983, 1986, and 1998, respectively.

In 1986, he joined the Department of Electrical Measurement and Instrumentation, China Institute of Metrology where he became a Lecturer in 1988 and an Associate Professor in 1997. After obtaining his Ph.D., he joined the Institute of Mechanics, Chinese Academy of Sciences as a postdoctoral researcher studying gas/liquid two-phase flows in horizontal pipes. Since 2007, he has been with Cranfield University, where he was a Research Fellow in the School of Engineering, and became a Senior Research Fellow in 2012. Currently he is a Senior Research Fellow in the Centre for Thermal Energy Systems and Materials. Prior to joining Cranfield University, he worked in the area of two-phase flows in closed channels in Imperial College London as a Research Associate for four years and a Research Fellow in the University of Cambridge for two years. He is a Member of Institute of Physics.

Dr Lao's current research interests cover process measurement & systems, and experimental fluid dynamics of multiphase flow systems. His expertise embraces both sensing technologies and multiphase fluid dynamics.

Gioia Falcone is Rankine Chair, Professor of Energy Engineering at the University of Glasgow. Until June 2018, she was Professor and Head of the Geo-Energy Engineering Centre at Cranfield University.

Between 2011 and early 2016, she held the Endowed Chair and Professorship in Geothermal Energy Systems at Clausthal University of Technology, Germany, where she was also the Director of the Institute of Petroleum Engineering. She was formerly an assistant and then associate professor in petroleum engineering at Texas A&M University, USA. Prior to joining academia, she worked with Eni-Agip, Enterprise Oil UK, Shell E&P UK and Total E&P UK.

She holds a *Laurea Summa Cum Laude* in environmental-petroleum engineering from Sapienza University of Rome, a M.Sc. degree in petroleum engineering from Imperial College London and a Ph.D. in chemical engineering from Imperial College London. Her research interests include multiphase flow and production optimisation.